Probabilistic concepts in formal contexts*

Alexander Demin[1], Denis Ponomaryov[1], Evgeny Vityaev[2]

[1] Institute of Informatics Systems, Novosibirsk, Russia
[2] Institute of Mathematics, Novosibirsk, Russia

alexandredemin@yandex.ru, ponom@iis.nsk.su, vityaev@math.nsc.ru

Abstract. We generalize the main notions of Formal Concept Analysis with the ideas of the semantic probabilistic inference. We demonstrate that under standard restrictions, our definitions completely correspond to the original notions of Formal Concept Analysis. From the point of view of applications, we propose a method of recovering concepts in formal contexts in presence of noise on data.

1 Introduction

Assume that a scientist needs to classify some finite set of objects with respect to n attributes. The objects are observed in a number of experiments, where each of them is assigned a certain set of attributes. The results of each experiment can be represented as a table, with rows labelled by the object names, columns labelled by the attribute names, and each cell (i, j) filled iff object i has attribute j. Having results of one particular experiment it is reasonable to classify the objects in the following way: put those objects in groups which have a common set of attributes and no object out of this group has these attributes. It is well known that the pairs $\langle$object set, attribute set$\rangle$ of this kind can be naturally ordered and represented in a convenient way as studied in Formal Concept Analysis [2, 3] (FCA). Now assume that we know the results of the whole body of the experiments and we would like to build a classification of the objects with respect to the whole collection of data. Typically, an object may have some attribute in a number of experiments and lack this attribute in the remaining number of them. To cope with this ambiguity when building classifications, we employ the method of semantic probabilistic inference introduced in [10–12]. In this paper, we generalize the standard notion of truth of an implication on data by means of a truth valuation based on a probability measure. We define an analog of the classification unit studied in FCA in terms of fixed points of implications which hold on data with respect to this valuation. To the best of our knowledge, there are no published papers describing similar probabilistic approaches. For instance, in [1] the main notions of Formal Concept Analysis are reformulated in terms

* This work was funded by the grant of the President of Russian Federation (grant No. MK-2037.2011.9), the Russian Foundation for Basic Research (grant No. 11-07-00560a), and Integration Projects of the Siberian Division of the Russian Academy of Sciences (grant No. 47, 111, 119).

of a probability logic, but their definitions are not generalized in the scope of FCA. By this work, we aim at establishing connections between Formal Concept Analysis and the method of semantic probabilistic inference. The contribution of this paper is the generalization of the key notions of FCA in terms of the semantic probabilistic inference.

2 Preliminaries

Let us start with basic definitions and results from Formal Concept Analysis.

Definition 1. *A formal context is a triple* (G, M, I)*, where* G *and* M *are sets and* $I \subseteq G \times M$ *is a relation between the elements of* G *and* M*. The elements of* G *are called objects and the elements of* M *are called attributes of the context. We call a context finite if* G *and* M *are finite sets.*

For brevity, we omit the word "formal" and call the triples (G, M, I) from the definition above *contexts*. Every context can be naturally represented in a tabular form, as noted in the introduction. For a context (G, M, I), we define the operation $'$ on the subsets $A \subseteq G$ and $B \subseteq M$ as follows:

$$A' = \{m \in M \mid \forall\, g \in A\ (g, m) \in I\}, \quad B' = \{g \in G \mid \forall\, m \in B\ (g, m) \in I\}.$$

For $g \in G$, the set $\{g\}'$ will be abbreviated by the notation g'.

Definition 2. *A concept in context* (G, M, I) *is a pair* (A, B) *with* $A \subseteq G,\ B \subseteq M,\ A' = B$*, and* $B' = A$*. The set* A *is called the extent and* B *the intent of the concept* (A, B)*.*

In fact, a *concept* can be viewed as the classification unit which groups objects and attributes of a context.

The following simple fact will is frequently used in proofs of the main claims in this paper:

Lemma 1. *If* (G, M, I) *is a context and* $B_1, B_2 \subseteq M$ *are sets of attributes, then*

1. $B_1 \subseteq B_2 \Longrightarrow B_2' \subseteq B_1'$
2. $B_1 \subseteq B_1''$.

Definition 3. *A (partial) order* $\leqslant$ *on concepts is defined as follows: if* (A_1, B_1) *and* (A_2, B_2) *are concepts of a context then* $(A_1, B_1) \leqslant (A_2, B_2)$ *if* $A_1 \subseteq A_2$ *(or, equivalently by Lemma 1, if* $B_2 \subseteq B_1$*).*

Theorem. *The relation* $\leqslant$ *induces a complete lattice on the set of concepts of a context, with the infimum and supremum of subsets given, respectively, by:*

$$\bigwedge_{j \in J} (A_j, B_j) = (\underset{j \in J}{\cap} A_j,\ (\underset{j \in J}{\cup} B_j)'')$$

$$\bigvee_{j \in J} (A_j, B_j) = ((\underset{j \in J}{\cup} A_j)'',\ \underset{j \in J}{\cap} B_j).$$

Example 1. *Consider the finite context $K = (\{g_1, g_2, g_3, g_4\},\ \{m_1, m_2, m_3, m_4\}, I)$ represented in the tabular form in Figure 1. The lattice of all concepts in context K is given in the same Figure; each element of the lattice is labelled by the set of objects and the set of attributes which are, respectively, the extent and the intent of the corresponding concept.*

I	m_1	m_2	m_3	m_4
g_1	×	×	×	
g_2	×	×		
g_3			×	
g_4				×

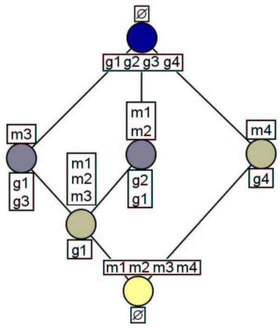

Fig. 1. A context and the corresponding concept lattice.

The procedures of computing the complete concept lattice for a given finite context [7, 8] are one of the basic algorithms in Formal Concept Analysis. In fact, they provide a classification of objects of a context with respect to their attributes and allow for finding all possible classes.

If $K = (G, M, I)$ is a context, we may speak about the truth of the following statements on K: "all objects having attribute set $B_1 \subseteq M$ also have attributes $B_2 \subseteq M$". As all properties of a context are in some sense symmetric with respect to the sets G and M, we can formulate the similar statements about subsets of G: "all attributes having the set $A_1 \subseteq G$ as their objects also have the set $A_2 \subseteq G$". W.l.o.g. we consider the statements only of the first kind. In fact, they define a monotone operator, an implication, on the boolean algebra of subsets of M. If a context K is finite, then clearly the set of all such statements true on K is also finite. Let us formalize the notion of an implication true on a context by the definitions from Chapter 2.3 in [2].

Definition 4. *An implication on a set M is an ordered pair of subsets $A, B \subseteq M$ denoted as $A \to B$. The set A is called the premise and B the conclusion of the implication $A \to B$. A subset $T \subseteq M$ respects an implication $A \to B$ if $A \not\subseteq T$ or $B \subseteq T$. A family of subsets of M respects an implication $A \to B$ if every set of this family respects $A \to B$.*

An implication $A \to B$ holds on a context $K = (G, M, I)$ (notation $K \models A \to B$) if $A, B \subseteq M$ and the family of sets $\{g' \mid g \in G\}$ respects $A \to B$.

The premise of an implication $A \to B$ is said to be false on a context $K = (G, M, I)$ if there is no element $g \in G$ such that $A \subseteq g'$. An implication $A \to B$ is called a tautology if $B \subseteq A$.

For a context $K = (G, M, I)$, we denote by $Imp(K)$ the set of all implications on M which hold on K. It is easy to verify that the tautologies and the set of implications whose premise is false in K are subsets of $Imp(K)$. When ambiguity does not arise, we will use the same symbol $\models$ to denote that a set or a family of sets respects an implication.

Every family L of implications on a set M defines the monotone operator $f_L : 2^M \to 2^M$ given by

$$f_L(X) = X \cup \{B \mid A \to B \in L,\ A \subseteq X\}.$$

Clearly, for each $X \subseteq M$, it holds $f_L(X) = X \Leftrightarrow X \models L$.

Remark 1. *Let L be a family of implications on a set M. Then for each $X \subseteq M$, there exists a minimal set $Y \subseteq M$ such that $X \subseteq Y$ and $f_L(Y) = Y$.*

Therefore, any family L of implications on a set M defines the operator $\bar{f}_L : 2^M \to 2^M$ which for every $X \subseteq M$ gives the minimal subset $Y \subseteq M$ satisfying the conditions of the remark. Clearly, for each $X \subseteq M$, we have $f_L(X) = X \Leftrightarrow \bar{f}_L(X) = X$.

Remark 2. *If $K = (G, M, I)$ is a context and $A \to B$ is an implication on M then $K \models A \to B \Leftrightarrow \forall m \in B\ (K \models A \to \{m\})$.*

In the following, we consider implications only of the form $A \to \{m\}$ and use the notation $A \to m$ for them.

If K is a context, then for every implication $A \to m \in Imp(K)$, there exists a set $\{A_0 \to m \in Imp(K) \mid A_0 \subseteq A$ and for each $A_1 \subseteq A$, if $A_1 \subset A_0$ then $A_1 \to m \notin Imp(K)\}$. For a context K, let us denote by $MinImp(K)$ the set of all implications of the form $A_0 \to m \in Imp(K)$ in which the premise A_0 is minimal in the above mentioned sense. We note that this definition is a variant of the notion of a law in [10–12].

Below, we give a slightly modified formulation of Proposition 20 from [2] which is central for results in this paper.

Proposition 1. *Let $K = (G, M, I)$ be a context, $T \subseteq Imp(K)$ be the set of tautologies on M, and $F \subseteq Imp(K)$ be the set of implications whose premise is false on K. Then for every subset $B \subseteq M$, we have:*

1. *$f_{MinImp(K) \setminus T}(B) = B \Leftrightarrow B'' = B$;*
2. *if $B' \neq \varnothing$ then $f_{MinImp(K) \setminus \{F \cup T\}}(B) = B \Leftrightarrow B'' = B$.*

It is straightforward by Definition 2 that for each context $K = (G, M, I)$, a subset $B \subseteq M$ is an intent of some concept in context K iff $B'' = B$. Therefore, as soon as a context $K = (G, M, I)$ is given, we have the set $Imp(K)$ of all implications which hold on K and the fixed points of the operator $f_{MinImp(K)\setminus T}$: $2^M \to 2^M$ correspond exactly to the intents of the concepts of K. If we omit the set F of implications from $MinImp(K) \setminus T$ whose premise is false on K then the fixed points of $f_{MinImp(K)\setminus\{F\cup T\}} : 2^M \to 2^M$ correspond to the intents of the concepts of K excluding the single concept $(\varnothing, M)$. Because for each $B \subseteq M$, the condition $B'' \neq M$ obviously yields $B' \neq \varnothing$.

3 Probabilistic concepts on classes of contexts

In Section 2, we have defined the notion of truth of an implication on a given context. Let us now demonstrate how this notion can be generalized with a truth valuation wrt a class of contexts. In this section, we proceed to ideas of the method of semantic probabilistic inference in application to FCA. As described in [10–12], the regularities on data (in particular, implications) are formalized in this method as universal formulas of the first order language of a countable signature consisting of predicates and constants. Thus, the standard notion of implication defined in [2] is far more specific than the concept of regularity on data considered in the semantic probabilistic inference (we note that there have been studied implications in papers on FCA which also go far beyond the definitions in [2]). However, in order to show this method useful in the case of FCA, it will be convenient to stay within the standard algebraic definitions. For this reason, we further present some restriction of the method of semantic probabilistic inference in terms common in Formal Concept Analysis.

Definition 5. *A class of contexts over sets G and M is a family $\mathcal{K} = \{(G, M, I_j)\}_{j\in J\neq\varnothing}$, where for each $j \in J$, the triple (G, M, I_j) is a context. We use the notation $\mathcal{K}(G, M)$ for a class $\mathcal{K}$ of contexts over sets G and M. A probability model of type I is a pair $\mathcal{M} = (\mathcal{K}(G, M), \rho)$, where $G \neq \varnothing$ and ρ is a probability measure on the set $\mathcal{K}$ satisfying the condition:* $\forall\ S_1, S_2 \subseteq G\times M\ \ \forall\ (G, M, I) \in \mathcal{K}$

$$S_1 \not\subseteq I \text{ or } S_2 \subseteq I \iff \rho(\{(G, M, I_j) \mid S_1 \cup S_2 \subseteq I_j\}) = \rho(\{(G, M, I_j) \mid S_1 \subseteq I_j\}).$$

For a subset $S \subseteq G \times M$, we call the value of the function $\nu_{\mathcal{M}}(S) = \rho(\{(G, M, I) \in \mathcal{K} \mid S \subseteq I\})$ the probability of the set S on $\mathcal{M}$.

For brevity, in this section we call the pair $(\mathcal{K}(G, M), \rho)$ from the definition above the *probability model* or simply, *model*.

Let $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ be a probability model and $A \to m$ be an implication on the set M. An instantiation of $A \to m$ on the model $\mathcal{M}$ is a pair $\langle g, A \to m\rangle$, where $g \in G$. The value of the function

$$\mu_{\mathcal{M}}(\langle g, A \to m\rangle) = \begin{cases} \frac{\nu_{\mathcal{M}}(\ S\cup\{<g,m>\}\ )}{\nu_{\mathcal{M}}(S)} & \text{if } \nu_{\mathcal{M}}(S) \neq 0,\ \text{ where } S = \{< g, a >\mid a \in A\} \\ \text{undefined}, & \text{otherwise} \end{cases}$$

is called the probability of the instantiation $\langle g, A \to m\rangle$ on the model $\mathcal{M}$.

If $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ is a probability model and $A \to m$ is an implication on the set M then the value of the function

$$\eta_{\mathcal{M}}(A \to m) = \begin{cases} \text{undefined} & \text{if } \forall g \in G \ \mu_{\mathcal{M}}(\langle g, A \to m \rangle) \text{ is undefined} \\ \inf_{g \in G} \mu_{\mathcal{M}}(\langle g, A \to m \rangle), \text{ otherwise} \end{cases}$$

is called the probability of the implication $A \to m$ on the model $\mathcal{M}$.

Remark 3. *Let $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ be a probability model and $A \to m$ be an implication on the set M whose probability is defined on $\mathcal{M}$. Then $\eta_{\mathcal{M}}(A \to m) = 1$ iff $\forall K \in \mathcal{K}$ $(A \to m \in Imp(K))$.*

Definition 6. *Let $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ be a probability model and $imp(M)$ be the set of all those implications on M whose probability is defined on $\mathcal{M}$. We call an implication $A \to m \in imp(M)$ a probabilistic law on $\mathcal{M}$, if the following conditions hold:*

- *$\eta_{\mathcal{M}}(A \to m) \neq 0$;*
- *if $A_0 \to m \in imp(M)$ and $A_0 \subset A$ then $\eta_{\mathcal{M}}(A_0 \to m) < \eta_{\mathcal{M}}(A \to m)$.*

An implication $A \to m \in imp(M)$ is called a maximally specific probabilistic law on $\mathcal{M}$ if it is a probabilistic law on $\mathcal{M}$, $A \neq \{m\}$, and there is no probabilistic law $A_0 \to m$ on $\mathcal{M}$ such that $A \subset A_0$ and $A_0 \to m$ is not a tautology.

Remark 4. *If an implication is a maximally specific probabilistic law on $\mathcal{M}$ then it is not a tautology.*

Definition 7. *Let $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ be a probability model and $S(\mathcal{M})$ be the set of all maximally specific probabilistic laws on $\mathcal{M}$. An implication $A \to m \in S(\mathcal{M})$ is called the strongest probabilistic law on $\mathcal{M}$ if its probability value on $\mathcal{M}$ is maximal among all implications $B \to m \in S(\mathcal{M})$. We use the notation $D(\mathcal{M})$ for the set of all strongest probabilistic laws on $\mathcal{M}$.*

Due to the the minor restrictions on the function ρ in the definition of the probability model, the existence of the maximum in the sense of Definition 7 is not guaranteed. Thus, the existence of the strongest probabilistic laws is not guaranteed either. However, we describe further in this section a way to define a probability model (based on a finite class of finite contexts), which gives a large class of models guaranteeing the existence of such implications. Note that in general for a given m, there may exist several strongest probabilistic laws of the form $A \to m$.

Informally, every implication on a probability model can be seen as a "prediction" (wrt some measure of truth) of the fact that each object having the attributes from the premise will also have the attribute from the conclusion. Similarly to Formal Concept Analysis (recall Proposition 1), implications in the method of semantic probabilistic inference are directly related to the process of grouping objects and attributes into classification units. If data are represented by a class $\mathcal{K}$ of contexts then the choice of implications wrt their probability on a model $(\mathcal{K}, \rho)$ becomes central for generating classes on the basis of the provided data. The definition of a minimal implication (as given by the set $MinImp(K)$), probabilistic law, maximally specific and the strongest probabilistic law are adopted from the corresponding definitions in [10–12, 15] to the case of FCA. Such implications have a number of useful theoretical and practical properties which justify their application:

- the set of all minimal implications which hold on every context from a class $\mathcal{K}$ gives, in some sense, an axiomatization of this class of contexts: the implicational theory of $\mathcal{K}$ (restricted to implications with non-false premises) semantically follows from it [10, 12] (the analogue of the Duquenne-Guigues theorem on implication base [4]);
- a probabilistic law excludes the possibility that an attribute in the conclusion can be "predicted" by a proper subset of the premise with probability greater than the probability of the law itself; together with the requirement of maximal specificity, this leads in practice to grouping attributes into smaller classes, with greater probability [9];
- it is proved in [10, 11] that in case negative information is allowed in implications, the set of maximally specific probabilistic laws is consistent (i.e. there can not be a situation when the presence and absence of some attribute are "predicted" simultaneously);
- the strongest probabilistic laws lead to assigning an attribute to a class which "predicts" it with maximal probability; at the same time, this does not rule out situations when the same attribute can belong to different classes [14];
- the **Discovery** software tool is implemented which allows to find the above mentioned types of implications on tabular data and compute the corresponding object-attribute classes; this software has proved successful in a large number of applications [5, 10, 15].

Definition 8. *Let $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ be a probability model of type I. A pair of sets (A, B) is called a probabilistic concept of a context $(G, M, I) \in \mathcal{K}$ in model $\mathcal{M}$ if it satisfies the following conditions:*

- $A \subseteq G$, $B \subseteq M$,
- $f_{D(\mathcal{M})}(B) = B$,
- $\exists E \subseteq B$ $(\bar{f}_{D(\mathcal{M})}(E) = B$ *and* $E \neq \varnothing \neq E')$,
- $A = \bigcup\{E' \mid \varnothing \neq E \subseteq B,\ \bar{f}_{D(\mathcal{M})}(E) = B\}$,

where $'$ is the operation in the context (G, M, I). The set A is called the extent and B the intent of the probabilistic concept (A, B).

Therefore, given a probability model $\mathcal{M} = (\mathcal{K}(G, M), \rho)$, the set of the fixed points of the operator $f_{D(\mathcal{M})}$ restricts the set of all possible probabilistic concepts of contexts from the class $\mathcal{K}$ in the model $\mathcal{M}$.

Theorem 1. *Consider a context $K = (\varnothing \neq G, M, I)$ and a probability model $\mathcal{M} = (\{K\}, \rho)$. For all non-empty subsets $A \subseteq G$ and $B \subseteq M$, the pair (A, B) is a concept in context K iff (A, B) is a probabilistic concept of context K in model $\mathcal{M}$.*

Let $\mathcal{K} = \{(\varnothing \neq G, M, I_j)\}_{j \in J \neq \varnothing}$ be a finite class consisting of finite contexts. We now describe a natural way to define a probability model $(\mathcal{K}, \rho)$ on the class $\mathcal{K}$. For each context $K \in \mathcal{K}$, we set $\rho(\{K\}) = 1/|J|$ and for a subset $\mathcal{C} \subseteq \mathcal{K}$ define $\rho(\mathcal{C}) = \sum_{K \in \mathcal{C}} \rho(\{K\})$.

Then ρ is a discrete probability measure on $\mathcal{K}$ and for every $S \subseteq G \times M$, we have $\nu_{\mathcal{M}}(S) = |\tilde{J}|/|J|$, where $\tilde{J}$ is the maximal subset of J satisfying the condition $\forall j \in \tilde{J}\ (S \subseteq I_j)$. It is easy to verify that $(\mathcal{K}, \rho)$ is indeed, a probability model. We call a model defined in this way the *frequency probability model* (of type I).

Let us illustrate the given definitions.

Example 2. *Consider the sets $G = \{g_1, g_2\}$, $M = \{m_1, m_2, m_3\}$, and the class $\mathcal{K} = \{(G, M, I_j)\}_{j \in \{1,2,3\}}$ consisting of three contexts given in the tabular form below.*

I_1	m_1	m_2	m_3
g_1	×		×
g_2		×	

I_2	m_1	m_2	m_3
g_1		×	×
g_2	×	×	

I_3	m_1	m_2	m_3
g_1	×		×
g_2	×	×	

Then the pairs $(\{g_1\}, \{m_1, m_2, m_3\})$ and $(\{g_1, g_2\}, \{m_1, m_2\})$ are the only probabilistic concepts of the context (G, M, I_1) in the frequency probability model $\mathcal{M} = (\mathcal{K}, \rho)$.

Proof. The probability measure ρ defines uniquely the value $\eta_{\mathcal{M}}(A \to m)$ for each implication $A \to m$ on the set M. In the tables below, we give the probability of each implication of the form $A \to m$ on $\mathcal{M}$ which is not a tautology.

$A \to m$	$\eta_{\mathcal{M}}(A \to m)$
$\{\varnothing\} \to m_1$	2/3
$m_2 \to m_1$	0
$m_3 \to m_1$	2/3
$m_2, m_3 \to m_1$	0
$\{\varnothing\} \to m_2$	1/3
$m_1 \to m_2$	0

$A \to m$	$\eta_{\mathcal{M}}(A \to m)$
$m_3 \to m_2$	1/3
$m_1, m_3 \to m_2$	0
$\varnothing \to m_3$	0
$m_1 \to m_3$	0
$m_2 \to m_3$	0
$m_1, m_2 \to m_3$	0

The premises of those implications which form the set $D(\mathcal{M})$ of the strongest probabilistic laws on $\mathcal{M}$ are written in parentheses. Let us give an example of computing the probability of one of the implications from the table above: $\eta_{\mathcal{M}}(m_3 \to m_1) = inf_{g \in G}\ \mu_{\mathcal{M}}(\langle g, m_3 \to m_1 \rangle) = inf_{g \in G} \frac{\nu_{\mathcal{M}}(\{<g,m_3>,<g,m_1>\})}{\nu_{\mathcal{M}}(\{<g,m_3>\})} = \frac{\nu_{\mathcal{M}}(\{<g_1,m_3>,<g_1,m_1>\})}{\nu_{\mathcal{M}}(\{<g_1,m_3>\})} = 2/3$, because the value of $\mu_{\mathcal{M}}(\langle g_2,\ m_3 \to m_1 \rangle)$ is undefined due to $\nu_{\mathcal{M}}(\{< g_2, m_3 >\}) = 0$. Note that the implication $m_3 \to m_1$ is not a probabilistic law, because there exists the implication $\varnothing \to m_1$ having the same probability on $\mathcal{M}$.

Let us give the values of the operator $f_{D(\mathcal{M})}$ on the subsets $B \subseteq M$:

$B \subseteq M$	$f_{D(\mathcal{M})}(B)$
m_1	m_1, m_2
m_2	m_1, m_2
m_3	m_1, m_2, m_3
m_1, m_2	m_1, m_2

$B \subseteq M$	$f_{D(\mathcal{M})}(B)$
m_1, m_3	m_1, m_2, m_3
m_2, m_3	m_1, m_2, m_3
m_1, m_2, m_3	m_1, m_2, m_3
$\varnothing$	m_1, m_2

Exactly two subsets $B \subseteq M$ satisfy the condition $f_{D(\mathcal{M})}(B) = B$, namely the sets $\{m_1, m_2\}$ and $\{m_1, m_2, m_3\}$. Finally, we have:

$$\bigcup\{E' \mid \varnothing \neq E \subseteq \{m_1, m_2\},\ \bar{f}_{D(\mathcal{M})}(E) = \{m_1, m_2\}\} = \{g_1, g_2\},$$

$$\bigcup\{E' \mid \varnothing \neq E \subseteq \{m_1, m_2, m_3\},\ \bar{f}_{D(\mathcal{M})}(E) = \{m_1, m_2, m_3\}\} = \{g_1\}.$$

The single subset $E \subseteq \{m_1, m_2, m_3\}$ satisfying the conditions in the definition of a probabilistic concept is the set $\{m_3\}$ for which we have $\{m_3\}' = g_1$.

Therefore, we conclude that $(\{g_1\}, \{m_1, m_2, m_3\})$ and $(\{g_1, g_2\}, \{m_1, m_2\})$ are the only probabilistic concepts of context (G, M, I_1) in model $\mathcal{M}$. □

4 Probabilistic concepts on one context

In Section 3, we have considered the notion of a probability model of type I defined on a class of contexts. In fact, every class $\mathcal{K}$ of contexts which allows to define a probability measure, raises a set of probability models and thus defines possible families of the strongest probabilistic laws. Based on such implications, we made a "prediction" of the existence of attributes for objects in an arbitrary chosen context from class $\mathcal{K}$. Similarly to this approach, we may define the strongest probabilistic laws on the basis of only one given context. For this, we need only to slightly modify Definition 5 of a probability model.

Definition 9. *A probability model of type II (a probabilistic context) is a pair $\mathcal{M} = (K, \rho)$, where $K = (G, M, I)$ is a context and ρ is a probability measure on the set G satisfying the condition*

$$\forall\ B, C \subseteq M\ (B' \subseteq C' \Leftrightarrow \rho((B \cup C)') = \rho(B')).$$

If $B \to m$ is an implication on the set M then the value of the function

$$\eta_{\mathcal{M}}(B \to m) = \begin{cases} \frac{\rho((B \cup \{m\})')}{\rho(B')} & \textit{if } \rho(B') \neq 0 \\ \textit{undefined}, & \textit{otherwise} \end{cases}$$

is called the probability of $B \to m$ on the model $\mathcal{M}$.

For brevity, in this section we call the pair (K, ρ) from the definition above the *probability model* or simply, *model* and use the notation $K(G, M)$ for a context K over the set of objects G and the set of attributes M.

For a finite context $K = (\varnothing \neq G, M, I)$, a model $\mathcal{M} = (K, \rho)$ is called a *frequency probability model* (of type II) if the function ρ is defined as $\rho(\{g\}) =$

$1/|G|$ for every $g \in G$ and $\rho(A) = \sum_{g\in A} \rho(\{g\})$ for each subset $A \subseteq G$. We have $\forall B \subseteq M$ $(\rho(B') = |B'|/|G|)$. Note that $\mathcal{M}$ is indeed, a model, since for all subsets $B, C \subseteq M$ it holds that $B' \subseteq C' \Leftrightarrow (B \cup C)' = B' \Leftrightarrow |(B \cup C)'| = |B'|$.

Remark 5. *If $\mathcal{M} = (K(G, M), \rho)$ is a probability model and $B \to m$ is an implication on M then $\eta_{\mathcal{M}}(B \to m) = 1$ iff $B \to m \in Imp(K)$ and $B' \neq \varnothing$ (where $'$ is the operation in the context K).*

Let us define the notions of a probabilistic law, maximally specific probabilistic law, and the strongest probabilistic law on a model of type II in full accordance with Definitions 6 and 7. In the following, we use the same notation $D(\mathcal{M})$ as in Section 3 for the set of all strongest probabilistic laws on a model $\mathcal{M}$ of type II.

Proposition 2. *Let $\mathcal{M} = (K(G, M), \rho)$ be a probability model and $S \subseteq Imp(K)$ be the set of all tautologies on M and all the implications whose premise is false on K. Then we have $MinImp(K) \setminus S \subseteq D(\mathcal{M})$.*

Definition 10. *Let $\mathcal{M} = (K(G, M), \rho)$ be a probability model of type II. A pair of sets (A, B) is called a probabilistic concept in model $\mathcal{M}$ (a concept in probabilistic context $\mathcal{M}$) if it satisfies the conditions of Definition 8.*

Let $\mathcal{M} = (K, \rho)$ be a model with $K = (G, M, I)$. Consider the context $\overline{K} = (G, M, \bar{I})$, where $\bar{I} = \{< g, m > \mid \ g \in G,\ m \in \bar{f}_{D(\mathcal{M})}(g')\}$, $'$ is the operation in the context K. In other words, we have $I \subseteq \bar{I}$ and the relation $\bar{I}$ is obtained from I by adding the pairs $< g, m >$ "predicted" by the implications in $D(\mathcal{M})$. To clarify the connection between the concepts in context K and probabilistic concepts in model $\mathcal{M}$, we need to note that *the following statement is false in both directions:*

for all non-empty subsets $A \subseteq G$ and $B \subseteq M$, the pair (A, B) is a probabilistic concept in the model $\mathcal{M}$ iff (A, B) is a concept in the context $\overline{K}$.

To prove this, it is sufficient to consider any of the contexts $K_1 = (\{g_1, g_2\}, \{m_1\},\ I_1)$, $K_2 = (\{g_1, g_2\}, \{m_1, m_2, m_3\},\ I_2)$ given below together with the corresponding frequency probability models $\mathcal{M}_1 = (K_1, \rho_1)$ and $\mathcal{M}_2 = (K_2, \rho_2)$.

I_1	m_1
g_1	×
g_2	

I_2	m_1	m_2	m_3
g_1	×		
g_2		×	×

For these models, we have $D(\mathcal{M}_1) = \{\varnothing \to m_1\}$ and $D(\mathcal{M}_2) = \{\varnothing \to m_1,\ \{m_2\} \to m_3, \{m_3\} \to m_2\}$. Therefore, the set of all probabilistic concepts in the model $\mathcal{M}_1$ consists of the single concept $(\{g_1\}, \{m_1\})$ and the set $\{\ (\{g_1\}, \{m_1\}), (\{g_2\}, \{m_1, m_2, m_3\})\ \}$ represents all the probabilistic concepts in the model $\mathcal{M}_2$.

It is easy to check that for every $j = 1, 2$, the context $\overline{K}_j$ is obtained from K_j by setting $\bar{I}_j = I_j \cup \{< g_2, m_1 >\}$. It remains to note that the set of all concepts in the context $\overline{K}_1$ consists of the single pair $(\{g_1, g_2\}, \{m_1\})$ and the set $\{\ (\{g_1, g_2\}, \{m_1\}),\ (\{g_2\}, \{m_1, m_2, m_3\})\ \}$ represents all the concepts in the context $\overline{K}_2$.

Nevertheless, the following property is guaranteed which characterizes the connection between concepts in a context K and probabilistic concepts in the model $\mathcal{M} = (K, \rho)$.

Theorem 2. *Every probability model $\mathcal{M} = (K(G, M), \rho)$ has the following properties:*

1. *for each concept (A, B) in context K with $A \neq \varnothing \neq B$, there exists a probabilistic concept (A_1, B_1) in model $\mathcal{M}$ such that $A \subseteq A_1$ and $B \subseteq B_1$;*
2. *if (A_1, B_1) is a probabilistic concept in model $\mathcal{M}$ then there exists a concept (A, B) in context K with $\varnothing \neq A \subseteq A_1$ and $\varnothing \neq B \subseteq B_1$. Moreover, the set A_1 is the union of the extents of some of these concepts.*

Below, we give schemas of computation procedures for finding the set of probabilistic laws and probabilistic concepts on a given frequency probability model $\mathcal{M} = (K, \rho)$, where $K = (G, M, I)$ and $\forall m \in M\ (\{m\}' \neq \varnothing)$.

Let $S \subseteq Imp(K)$ be the set of all tautologies on M and all the implications whose premise is false on the context K. For the given context K, the cardinality of $MinImp(K) \setminus S$ can be exponential in the value of $|G| \times |M|$. This follows from Theorem 1 in [6], where the construction of the corresponding context is given. By Proposition 2, we have $MinImp(K) \setminus S \subseteq D(\mathcal{M})$ and by definition, the set of all probabilistic laws on $\mathcal{M}$ contains $D(\mathcal{M})$. For this reason, the procedure for finding the set of probabilistic laws is based on a heuristic.

Let us introduce some auxiliary definitions. For an implication $A \to m$ on a set M, the *length* of $A \to m$ is the cardinality of the set A; we use the notation $len(A \to m)$. Call an implication $A_2 \to m$ a *specification* of an implication $A_1 \to m$ if $A_2 = A_1 \cup \{n\}$, where $n \in M \setminus A_1$. For a family L of implications, $Spec(L)$ will denote the set of all possible specifications of implications from L.

The computation procedure for finding probabilistic laws is based on the concepts of the semantic probabilistic inference. The main idea is to extend stepwise the premises of implications and check the conditions in the definition of a probabilistic law at each step. This implements a directed enumeration of implications which allows to considerably reduce the search space. The reduction is achieved due to the application of the following heuristic: when the length of the generated implications reaches a certain value (called the *base enumeration depth*), the specification is applied only to those implications which are probabilistic laws.

For simplicity, we give a schema of the computation procedure for finding probabilistic laws of the form $A \to m$ on the model $\mathcal{M}$ for a chosen attribute $m \in M$. Besides the mentioned probability model $\mathcal{M}$ and the element $m \in M$, the additional input parameter of the procedure is the value d of the base enumeration depth, with $1 \leqslant d \leqslant |M|$. The output of the procedure is the set of the probabilistic laws found on the model $\mathcal{M}$ with the element m in the conclusion.

At step $k = 0$, the set $imp(\mathcal{M})_{(k)}$ of implications is generated which consists of the single implication of zero length of the form $R = \varnothing \to m$. For the implication R, the conditions on a probabilistic law in the Definition 6 are verified. Denote the set of all probabilistic laws computed at step k of the computation procedure by $REG^{(k)}_{\mathcal{M}}(m)$. If R is a probabilistic law then $REG^{(0)}_{\mathcal{M}}(m) = \{R\}$. Else, we have $REG^{(0)}_{\mathcal{M}}(m) = \varnothing$ and the procedure returns the empty set. Indeed, in this case we have $\eta_{\mathcal{M}}(\varnothing \to m) = 0$ and, by the definition of the model $\mathcal{M}$, the probability of each implication of the form $B \to m$ is either undefined, or equals zero on $\mathcal{M}$. This means that no such implication can be a probabilistic law on $\mathcal{M}$.

At step $1 \leqslant k \leqslant d$, the set $imp(\mathcal{M})_{(k)}$ of specifications is computed for all implications obtained at the previous step whose probability is defined and not equal to zero or one: $imp(\mathcal{M})_{(k)} = Spec(\{R \mid R \in imp(\mathcal{M})_{(k-1)},\ 0 < \eta_{\mathcal{M}}(R) < 1\})$. Each implication in this set is of length k. For every implication from $imp(\mathcal{M})_{(k)}$ the conditions in the definition of a probabilistic law are verified and thus the set $REG^{(k)}_{\mathcal{M}}(m)$ is formed.

At step $d < k \leqslant |M|$, the set $imp(\mathcal{M})_{(k)}$ of specifications is computed for all implications obtained at the previous step having a probability less than 1: $imp(\mathcal{M})_{(k)} = Spec(\{R \mid R \in REG^{(k-1)}_{\mathcal{M}}(m),\ \eta_{\mathcal{M}}(R) < 1\})$. For each of the obtained implications the conditions in the definition of a probabilistic law are verified and thus the set $REG^{(k)}_{\mathcal{M}}(m)$ is formed. The execution of the computation procedure ends either on the step $k = |M|$, or in case at some step $d < k < |M|$ no probabilistic laws are obtained, i.e. when $REG^{(k)}_{\mathcal{M}}(m) = \varnothing$. The resulting set of the probabilistic laws for the attribute m returned by the procedure is the union $\bigcup_k REG^{(k)}_{\mathcal{M}}(m)$.

To select the strongest (wrt the input parameters) probabilistic laws from the set of the computed implications, it suffices to directly verify the conditions of Definition 7.

The steps $k \leqslant d$ of the procedure are called base enumeration steps and those for $k > d$ are called additional enumeration steps. As proved by experiments, the base enumeration depth of value $d \leqslant 3$ suffices in a large number of applications. In practice, the inequalities in Definition 6 are verified with respect to a statistical criterion (e.g., Fisher's exact test for contingency tables) which is applied with a user defined confidence level α.

Let L be a non-empty set of probabilistic laws on the model $\mathcal{M}$. Note that in case L is the output of the above given procedure for base enumeration depth $d = |M|$, we have $L = D(\mathcal{M})$.

Let us describe an iterative procedure for finding probabilistic concepts in model $\mathcal{M}$ with respect to the family L of implications.
At step $k = 1$ the following set is generated: $C^{(1)} = \{\bar{f}_L(A \cup \{m\}) \mid A \to m \in L\}$.

At step $k > 1$, in case $C^{(k-1)} = \varnothing$, the procedure returns the list of all the computed probabilistic concepts. Otherwise, for each $B \in C^{(k-1)}$, having the family of implications $L_B = \{A \to m \in L \mid A \subseteq B\}$, the set $A = \{g \in G \mid g' \cap B \neq \varnothing,\ f_{L_B}(g' \cap B) = B\}$ is computed. If $A \neq \varnothing$ then the pair (A, B) is added to the list of the computed probabilistic concepts. Further, the set $C^{(k)} = \{\bar{f}_L(B \cup C) \mid B, C \in C^{(k-1)},\ \bar{f}_L(B \cup C) \notin C^{(k-1)}\}$ is generated and the next iteration is executed. The description of the procedure is complete.

Example 3. Consider the contexts K_1 and K_2 given in Figure 2. The concepts in context K_1 having a non-empty extent and intent are the pairs $(\{g_1, \ldots, g_{20}\}, \{m_1, \ldots, m_5\})$ and $(\{g_{21}, \ldots, g_{40}\}, \{m_6, \ldots, m_{10}\})$. The context K_2 was obtained from K_1 by adding a random noise. The task was to recover the initial concepts in context K_2. With the given algorithms, the set of the strongest probabilistic laws on the frequency model $\mathcal{M} = (K_2, \rho)$ was computed; it consisted of 22 implications. The set of probabilistic concepts in model $\mathcal{M}$ turned out to be equal to the set of concepts in the initial context K_1 with non-empty extents and intents.

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
g_1	×	×	×	×	×					
g_2	×	×	×	×	×					
g_3	×	×	×	×	×					
g_4	×	×	×	×	×					
g_5	×	×	×	×	×					
g_6	×	×	×	×	×					
g_7	×	×	×	×	×					
g_8	×	×	×	×	×					
g_9	×	×	×	×	×					
g_{10}	×	×	×	×	×					
g_{11}	×	×	×	×	×					
g_{12}	×	×	×	×	×					
g_{13}	×	×	×	×	×					
g_{14}	×	×	×	×	×					
g_{15}	×	×	×	×	×					
g_{16}	×	×	×	×	×					
g_{17}	×	×	×	×	×					
g_{18}	×	×	×	×	×					
g_{19}	×	×	×	×	×					
g_{20}	×	×	×	×	×					
g_{21}						×	×	×	×	×
g_{22}						×	×	×	×	×
g_{23}						×	×	×	×	×
g_{24}						×	×	×	×	×
g_{25}						×	×	×	×	×
g_{26}						×	×	×	×	×
g_{27}						×	×	×	×	×
g_{28}						×	×	×	×	×
g_{29}						×	×	×	×	×
g_{30}						×	×	×	×	×
g_{31}						×	×	×	×	×
g_{32}						×	×	×	×	×
g_{33}						×	×	×	×	×
g_{34}						×	×	×	×	×
g_{35}						×	×	×	×	×
g_{36}						×	×	×	×	×
g_{37}						×	×	×	×	×
g_{38}						×	×	×	×	×
g_{39}						×	×	×	×	×
g_{40}						×	×	×	×	×

K_1

	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}
g_1	×	×	×	×	×	×				
g_2	×	×	×	×	×			×		
g_3	×	×	×	×	×					×
g_4	×	×	×	×	×	×				
g_5	×	×	×	×						
g_6	×	×	×	×	×	×				
g_7	×		×	×	×					
g_8	×	×	×	×	×			×		
g_9	×	×	×	×						
g_{10}	×	×		×	×					
g_{11}	×	×		×	×					
g_{12}	×		×	×	×					
g_{13}	×	×	×	×	×				×	
g_{14}	×	×	×	×	×				×	
g_{15}	×	×	×	×	×	×				
g_{16}	×	×	×	×	×					×
g_{17}	×	×	×	×						
g_{18}	×	×	×	×	×			×		
g_{19}	×	×	×	×	×				×	
g_{20}	×	×	×	×	×		×			
g_{21}							×	×	×	×
g_{22}						×	×	×	×	
g_{23}						×	×		×	×
g_{24}		×				×	×	×	×	×
g_{25}	×					×	×	×	×	×
g_{26}						×		×	×	×
g_{27}						×	×	×		×
g_{28}						×		×	×	×
g_{29}							×	×	×	×
g_{30}							×	×	×	×
g_{31}					×	×	×	×	×	×
g_{32}				×		×	×	×	×	×
g_{33}	×					×	×	×	×	×
g_{34}	×					×	×	×	×	×
g_{35}						×		×	×	×
g_{36}					×	×	×	×	×	×
g_{37}	×					×	×	×	×	×
g_{38}						×		×	×	×
g_{39}					×	×	×	×	×	×
g_{40}	×					×	×	×	×	×

K_2

Fig. 2. Recovering concepts in the presence of noise.

5 Conclusions

It easy to note from Definition 5 and 9 that the distinction between the probability model of type I and II is rather subtle. In particular, for each model $\mathcal{M}_2 = (K, \rho_2)$ of type II with $K = (\varnothing \neq G, M, I)$, it is possible to define a model $\mathcal{M}_1 = (\mathcal{K}, \rho_1)$ of type I such that $D(\mathcal{M}_1) = D(\mathcal{M}_2)$. It suffices to set $\mathcal{K} = \{K_g \mid g \in G,\ K_g = (\{h\}, M, I_g),\ I_g = \{< h, m >|< g, m >\in I\}\}$ and define $\forall\ \ \mathcal{C} \subseteq \mathcal{K}\ \rho_1(\mathcal{C}) = \rho_2(\{g \mid K_g \in \mathcal{C}\})$. Then for every implication $B \to m$ on M, we have $\eta_{\mathcal{M}_1}(B \to m) = \mu_{\mathcal{M}_1}(\langle h, B \to m\rangle) = \frac{\rho_1(\{K_g | \{<h,n>| n \in B \cup \{m\}\} \subseteq I_g\})}{\rho_1(\{K_g | \{<h,n>| n \in B\} \subseteq I_g\})} = \frac{\rho_2((B \cup \{m\})')}{\rho_2(B')}$ and thus $\eta_{\mathcal{M}_1}(B \to m) = \eta_{\mathcal{M}_2}(B \to m)$ which clearly, yields $D(\mathcal{M}_1) = D(\mathcal{M}_2)$. However, it is important in practice to make a distinction between the analysis of data represented by a class of contexts and analysis of data on the basis of a single given context. In the first case, we have a problem of classification of objects which are observed in a number of experiments in which every object is assigned a certain set of attributes. In the second case, the classification of objects is based on a single context which represents the whole body of experimental data on these objects. A context uniquely determines whether an object has a particular attribute and Formal Concept Analysis provides tools for building precise classification of objects on the basis of a given context. On the other hand, discovering probabilistic laws on a model over a given context allows to obtain classification units which are stable with respect to noise.

Example 3 demonstrates that the noise of a certain level does not change the set of concepts in a context, i.e. the set of concepts with a non-empty extent and intent in a given context is equal to the set of probabilistic concepts in a new context obtained by adding noise into the initial one. There exist types of noise (a formal definition is given in [10]) such that any level of a noise of this kind does not change the set of concepts in a context; such noise is called *concept preserving* [10]. This raises the problem of characterization of these types of noise.

In the definitions of implications and probabilistic laws in this paper, the notion of negation was not present. Due to this, the formulation of Theorem 2 appeared to be weaker than expected. This is because the negation was not present in the fundamentals of Formal Concept Analysis and we aimed at giving the most simple generalization of the basic notions of this method. The generalization of FCA according to the given ideas will allow to formalize the notions of “natural classification” and “idealization” as defined in [10, 13]. The semantic probabilistic inference which is central in the definitions of probabilistic concepts has been first introduced for first order logic and provides a method for discovering rather complicated regularities on data in comparison to those considered in this paper. Moreover, in the relational approach described in monographs [5, 10], it is argued that the formalization of regularities in the language of first order logic is essential for analyzing the whole body of information contained in data. Some examples of such regularities are given at the Web page [15] at

http://math.nsc.ru/AP/ScientificDiscovery/pages/Examples_of_rules.html

References

1. J. Deogun, L. Jiang, Y. Xie, V. Raghavan. Probability logic modeling of knowledge discovery in databases. *In Proc. ISMIS'2003*, pp. 402–407, 2003.
2. B. Ganter, R. Wille. Formal Concept Analysis: Mathematical Foundations. Springer Verlag, 1999.
3. B. Ganter, G. Stumme, R. Wille (ed.). Formal Concept Analysis: Foundations and Applications. Springer Verlag, 2005.
4. J.– L. Guigues, V. Duquenne. Families minimales d'implications informatives resultant d'un tableau de données binaires. *Math. Sci. Humaines*, 95, pp. 5–18, 1986.
5. B. Kovalerchuk, E. Vityaev. Data Mining in Finance: Advances in Relational and Hybrid methods. Kluwer, 2000.
6. S. Kuznetsov. On the intractability of computing the Duquenne–Guigues base. *J. UCS*, 10(8), pp. 927–933, 2004.
7. S. Kuznetsov, S. Obiedkov. Comparing performance of algorithms for generating concept lattices *J. Exp. Theor. Artif. Intell.*, 14(2-3), pp. 189-216, 2002.
8. D. Merwe, S. Obiedkov, D. Kourie. AddIntent: A New Incremental Algorithm for Constructing Concept Lattices. *In Proc. ICFCA'2004*, pp. 372–385, 2004.
9. S. Smerdov, E. Vityaev. Synthesis of logic, probability, and learning: Formalizing the concept of prediction. *Siberian Electronic Mathematical Reports*, 6, pp. 340-365, 2009 *(in Russian, with english abstract)*.
10. E. Vityaev. Knowledge Discovery. Computational Cognition. Models of Cognitive Processes. Novosibirsk State University, 293 p., 2006 *(in Russian)*
11. E. Vityaev. The logic of prediction. *In Proc. 9th Asian Logic Conference*, pp. 263–276, 2006.
12. E. Vityaev, B. Kovalerchuk. Empirical theories discovery based on the Measurement Theory. *Mind and Machine*, 14(4), pp. 551–573, 2005.
13. E. Vityaev, K. Lapardin, I. Khomicheva, A. Proskura. Transcription factor binding site recognition by regularity matrices based on the natural classification method. *Intelligent Data Analysis*, Special Issue "New Methods in Bioinformatics" ed. by E. Vityaev and N. Kolchanov, 12(5), pp. 495–512, 2008.
14. E. Vityaev, S. Smerdov. New definition of prediction without logical inference. *In Proc. IASTED Int. Conf. on Computational Intelligence (CI 2009)*, ed. by B. Kovalerchuk, pp. 48-54, 2009.
15. Scientific Discovery website: http://math.nsc.ru/AP/ScientificDiscovery/

6 Appendix

Proofs for Section 2

Remark 1. *Let L be a family of implications on a set M. Then for each $X \subseteq M$, there exists a minimal set $Y \subseteq M$ such that $X \subseteq Y$ and $f_L(Y) = Y$.*

Proof. Consider the following straightforward inductive process of building extensions of a set $X \subseteq M$. First, set $X_0 = X$. If a set X_i is constructed then we define $X_{i+1} = f_L(X_i)$. Finally, we take $Y = \bigcup_{i\in\omega} X_i$. □

Proposition 1. *Let $K = (G, M, I)$ be a context, $T \subseteq Imp(K)$ be the set of tautologies on M, and $F \subseteq Imp(K)$ be the set of implications whose premise is false on K. Then for every subset $B \subseteq M$, we have:*

1. $f_{MinImp(K)\setminus T}(B) = B \Leftrightarrow B'' = B$;
2. *if* $B' \neq \varnothing$ *then* $f_{MinImp(K)\setminus\{F\cup T\}}(B) = B \Leftrightarrow B'' = B$.

Proof. Let us first demonstrate that for every subset $B \subseteq M$, we have $f_{Imp(K)}(B) = B$ iff $f_{MinImp(K)}(B) = B$. If $f_{Imp(K)}(B) \supset B$ for some B then (with Remark 2) there is an implication $A \to m \in Imp(K)$ such that $A \subseteq B$, but $m \notin B$. Then there exists $A_0 \to m \in MinImp(K)$ with $A_0 \subseteq A$ and thus, $A_0 \subseteq B$, $m \notin B$, and $f_{MinImp(K)}(B) \supset B$, a contradiction. The reverse direction of the claim is obvious, since $MinImp(K) \subseteq Imp(K)$.

Similarly, it is not hard to verify that $f_{MinImp(K)\setminus L}(B) = B \Leftrightarrow f_{Imp(K)\setminus L}(B) = B$, where $L = T$ or $L = F \cup T$. This follows from the fact that for each implication $A \to m$ on M and every subset $A_0 \subseteq A$, the condition $A \to m \notin T$ yields $A_0 \to m \notin T$. On the other hand, by Lemma 1, it follows from $A' \neq \varnothing$ that $A_0' \neq \varnothing$. Therefore, we will prove the both claims of the proposition with respect to the set $Imp(K)$, instead of $MinImp(K)$.

1.$\Leftarrow$: Assume $B'' = B$, $A_1 \to A_2 \in Imp(K)\setminus T$, and $A_1 \subseteq B$. We demonstrate that $A_2 \subseteq B$. For each $g \in B'$, we have $g' \supseteq A_2$, because $g' \supseteq B'' = B$ by Lemma 1 and implication $A_1 \to A_2$ holds on K. Therefore, $\bigcap\{g' \mid g \in B'\} \supseteq A_2$. On the other hand, $\bigcap\{g' \mid g \in B'\} = B''$ and since $B'' = B$, we obtain $B \supseteq A_2$.

1.$\Rightarrow$: By Lemma 1, in any case we have $B'' \supseteq B$, so let us assume that $f_{Imp(K)\setminus T}(B) = B$, but $B'' \nsubseteq B$. Then $B \not\models B \to B'' \notin T$ and it is sufficient to demonstrate $B \to B'' \in Imp(K)$ to obtain contradiction.

a) If $B' = \varnothing$ then this obviously holds, since there is no $g \in G$ such that $B \subseteq g'$, i.e. the premise of the implication is false on K.

b) Let $B' \neq \varnothing$; we need to show that $\forall g \in G\ (B \subseteq g' \Rightarrow B'' \subseteq g')$. Clearly $\forall g \in G\ (B \subseteq g' \Leftrightarrow g \in B')$ and, by Lemma 1, we have $\forall g \in B'\ (B'' \subseteq g')$. Therefore, if $B \subseteq g'$ for some $g \in G$ then $B'' \subseteq g'$, i.e. $B \to B'' \in Imp(K)$. Moreover, $B \to B'' \in Imp(K) \setminus F$, because $B' \neq \varnothing$.

2. The sufficiency follows from the proof of claim 1, since the condition $f_{MinImp(K)\setminus T}(B) = B$ clearly, yields $f_{MinImp(K)\setminus\{F\cup T\}}(B) = B$. The necessity is proved by item b above. □

Proofs for Section 3

Remark 3. *Let $\mathcal{M} = (\mathcal{K}(G, M), \rho)$ be a probability model and $A \to m$ be an implication on the set M whose probability is defined on $\mathcal{M}$. Then $\eta_\mathcal{M}(A \to m) = 1$ iff $\forall K \in \mathcal{K}\ (A \to m \in Imp(K))$.*

Proof. $\Rightarrow$: The condition $\eta_\mathcal{M}(A \to m) = 1$ means that for each $g \in G$, the value of $\mu_\mathcal{M}(\langle g, A \to m\rangle)$ is either undefined, or equals 1. Thus, for each $g \in G$ and every context $K \in \mathcal{K}$, by the definitions of $\mu_\mathcal{M}$ and ρ, we have $A \not\subseteq g'$ or $m \in g'$. This means that $\forall K \in \mathcal{K}\ (A \to m \in Imp(K))$.

$\Leftarrow$: Assume that $\eta_\mathcal{M}(A \to m) < 1$. Then there is $g \in G$ such that the value of $\mu_\mathcal{M}(\langle g, A \to m\rangle)$ is defined and strictly less than 1. Then there exists a context $K \in \mathcal{K}$ in which $A \subseteq g'$ and $m \notin g'$, but this means that $A \to m \notin Imp(K)$. $\square$

Theorem 1. *Consider a context $K = (\varnothing \neq G, M, I)$ and a probability model $\mathcal{M} = (\{K\}, \rho)$. For all non-empty subsets $A \subseteq G$ and $B \subseteq M$, the pair (A, B) is a concept in context K iff (A, B) is a probabilistic concept of context K in model $\mathcal{M}$.*

Proof. Let $S \subseteq Imp(K)$ be the set of all tautologies on M and all the implications whose premise is false on the context K. We demonstrate that $MinImp(K) \setminus S = D(\mathcal{M})$.

$\subseteq$: Consider an arbitrary implication $A \to m \in MinImp(K) \setminus S$. By the definition of $\mathcal{M}$, for each subset $S \subseteq G \times M$, we have $\rho(S) = 0$ iff $S \not\subseteq I$. As the premise A is not false on K, we conclude that the probability of $A \to m$ is defined on $\mathcal{M}$ and, by Remark 3, we obtain $\eta_\mathcal{M}(A \to m) = 1$. Due to minimality of A, every implication $A_0 \to m$ with $A_0 \subset A$ does not hold on K. Besides, A_0 is not false on K, since A is not false on K. It follows from Remark 3 that $\eta_\mathcal{M}(A_0 \to m) = 0$ and thus, $A \to m$ is a probabilistic law on $\mathcal{M}$. Since $\eta_\mathcal{M}(A \to m) = 1$ and $m \notin A$, we conclude that $A \to m \in D(\mathcal{M})$.

$\supseteq$: By the definition of $\mathcal{M}$, we have $\forall S \subseteq G \times M\ \rho(S) \in \{0, 1\}$, hence, for every implication $A \to m \in D(\mathcal{M})$, by the definition of a probabilistic law, we obtain $\eta_\mathcal{M}(A \to m) = 1$. Then, due to the definition of $\mu_\mathcal{M}$, the premise A is not false on K and by Remark 3 and 4, we obtain that $A \to m \in Imp(K) \setminus S$. Assume there exists an implication $A_0 \to m \in Imp(K)$ such that $A_0 \subset A$. Then $A_0 \to m \in Imp(K) \setminus S$ and $\eta_\mathcal{M}(A_0 \to m) = 1$, but this contradicts the condition that $A \to m$ is a probabilistic law on $\mathcal{M}$; thus, $A \to m \in MinImp(K) \setminus S$.

Let (A, B) be a probabilistic concept of context K in model $\mathcal{M}$. To show that (A, B) is a concept in context K it is sufficient to verify that $A' = B$ and $B' = A$. Consider the set $\mathcal{C} = \{E \subseteq B \mid \bar{f}_{D(\mathcal{M})}(E) = B,\ E \neq \varnothing \neq E'\}$; it is non-empty by the definition of a probabilistic concept. For each $E \in \mathcal{C}$, due to $\bar{f}_{D(\mathcal{M})}(E) = B$ and the proved above, there exists an implication $E \to B \in Imp(K)$. Then $B' \neq \varnothing$ and since $f_{D(\mathcal{M})}(B) = B$, by point 2 of Proposition 1, we obtain $B'' = B$. Moreover, it follows from $E \to B \in Imp(K)$ that for each $g \in E'$ we have $g' \supseteq B$. This means that for every $g \in \bigcup\{E' \mid E \in \mathcal{C}\} = A$, we have $g' \supseteq B$ and thus, $A \subseteq B'$. On the other hand, for each $E \in \mathcal{C}$, the condition

$E \subseteq B$ yields $B' \subseteq E'$, hence, $B' \subseteq \bigcup\{E' \mid E \in \mathcal{C}\} = A$. Therefore, we have $A = B'$ which together with the condition $B'' = B$ gives $A' = B$.

Let (A, B) be a concept in context K where A and B are non-empty sets. We show that (A, B) is a probabilistic concept of context K in model $\mathcal{M}$. As $A \neq \varnothing$, $B' = A$, we have $B' \neq \varnothing$ and since $B'' = B$, by point 2 of Proposition 1 and the proved above, we obtain $f_{D(\mathcal{M})}(B) = B$. It remains to verify that $A = \bigcup\{E' \mid E \in \mathcal{C}\}$, where $\mathcal{C} = \{E \subseteq B \mid E \neq \varnothing,\ \bar{f}_{D(\mathcal{M})}(E) = B\}$, since clearly $B \in \mathcal{C}$. We have $\bigcup\{E' \mid E \in \mathcal{C}\} \supseteq B' = A$. On the other hand, if $g \in \bigcup\{E' \mid E \in \mathcal{C}\}$ then there exists $E \in \mathcal{C}$ such that $g \in E'$ and thus, $g' \supseteq E$. As $\bar{f}_{D(\mathcal{M})}(E) = B$, we have $E \to B \in Imp(K)$, hence $g' \supseteq B$ and $g \in B' = A$. Thus, all the conditions in the definition of a probabilistic concept are fulfilled. □

Proofs for Section 4

Remark 5. *If $\mathcal{M} = (K(G, M), \rho)$ is a probability model and $B \to m$ is an implication on M then $\eta_{\mathcal{M}}(B \to m) = 1$ iff $B \to m \in Imp(K)$ and $B' \neq \varnothing$ (where $'$ is the operation in the context K).*

Proof. If $\eta_{\mathcal{M}}(B \to m) = 1$ then $\rho(B') \neq \varnothing$ and thus $B' \neq \varnothing$, i.e. the premise B is not false on K. On the other hand, this condition means that $\rho((B \cup \{m\})') = \rho(B')$, hence, $B' \subseteq \{m\}'$ which is equivalent to $B \to m \in Imp(K)$. The same argument proves the claim in the reverse direction. □

Proposition 2. *Let $\mathcal{M} = (K(G, M), \rho)$ be a probability model and $S \subseteq Imp(K)$ be the set of all tautologies on M and all the implications whose premise is false on K. Then we have $MinImp(K) \setminus S \subseteq D(\mathcal{M})$.*

Proof. For each implication $B \to m \in MinImp(K) \setminus S$, it holds that $B' \neq \varnothing$, hence, by Remark 5, we obtain $\eta_{\mathcal{M}}(B \to m) = 1$. The condition of maximal probability for $B \to m$ is satisfied and obviously, there can not exist a probabilistic law $B_1 \to m$ on $\mathcal{M}$ with $B \subset B_1$. Besides, the implication $B \to m$ is itself a probabilistic law, because, by the condition $B \to m \in MinImp(K) \setminus S$ and Remark 5, for every subset $B_0 \subset B$ we have $\eta_{\mathcal{M}}(B_0 \to m) < 1$. Thus, all the conditions in the definition of the strongest probabilistic law are satisfied and $B \to m \in D(\mathcal{M})$. □

Theorem 2. *Every probability model $\mathcal{M} = (K(G, M), \rho)$ has the following properties:*

1. *for each concept (A, B) in context K with $A \neq \varnothing \neq B$, there exists a probabilistic concept (A_1, B_1) in model $\mathcal{M}$ such that $A \subseteq A_1$ and $B \subseteq B_1$;*
2. *if (A_1, B_1) is a probabilistic concept in model $\mathcal{M}$ then there exists a concept (A, B) in context K with $\varnothing \neq A \subseteq A_1$ and $\varnothing \neq B \subseteq B_1$. Moreover, the set A_1 is the union of the extents of some of these concepts.*

Proof. 1. Let $S \subseteq Imp(K)$ be the set of all tautologies on M and all the implications whose premise is false on K. As (A, B) is a concept in context K, we have $B'' = B$, $B' = A \neq \varnothing$ and by Proposition 1, we obtain $f_{MinImp(K)\setminus S}(B) = B$.

By Proposition 2, the following inclusion holds: $MinImp(K) \setminus S \subseteq D(\mathcal{M})$. Besides, for all families L_1 and L_2 of implications on M and any subset $B \subseteq M$, if $L_1 \subseteq L_2$ then $\bar{f}_{L_1}(B) \subseteq \bar{f}_{L_2}(B)$; therefore, we have $B \subseteq \bar{f}_{D(\mathcal{M})}(B)$. Denote $B_1 = \bar{f}_{D(\mathcal{M})}(B)$, $\mathcal{C} = \{E \subseteq B_1 \mid \bar{f}_{D(\mathcal{M})}(E) = B_1,\ E \neq \varnothing \neq E'\}$, and $A_1 = \cup\{E' \mid E \in \mathcal{C}\}$. Then obviously, $\bar{f}_{D(\mathcal{M})}(B_1) = B_1$. Note that $B \in \mathcal{C}$, $A = B'$, and $B' \subseteq A_1$, thus, we have $A \subseteq A_1$ and (A_1, B_1) is the required probabilistic concept in $\mathcal{M}$.

2. Consider the set $\mathcal{C} = \{E \subseteq B_1 \mid \bar{f}_{D(\mathcal{M})}(E) = B_1,\ E \neq \varnothing \neq E'\}$ and an arbitrary $E \in \mathcal{C}$. We have $MinImp(K) \setminus S \subseteq D(\mathcal{M})$, so $\bar{f}_{MinImp(K)\setminus S}(E) \subseteq B_1$. Denote $B = \bar{f}_{MinImp(K)\setminus S}(E)$; then clearly, $\bar{f}_{MinImp(K)\setminus S}(B) = B$. Besides, it follows from $E \neq \varnothing \neq E'$ that $B \neq \varnothing \neq B'$, hence, by Proposition 1, we obtain $B'' = B$. On the other hand, we have $E \subseteq B$, thus, $E' \supseteq B'$ and $A_1 = \cup\{E' \mid E \in \mathcal{C}\} \supseteq B'$. We conclude that (B', B) is the required concept in context K.

Note that the condition $B = \bar{f}_{MinImp(K)\setminus S}(E)$ yields $E \to B \in Imp(K)$ which is equivalent to $E' \subseteq B'$; therefore, we obtain $E' = B'$. Because of the arbitrary selection of the set $E \in \mathcal{C}$ and the condition $A_1 = \cup\{E' \mid E \in \mathcal{C}\}$, we conclude that A_1 is the union of the extents of some concepts (A, B) in context K with $\varnothing \neq B \subseteq B_1$. $\square$